

A Parameterized Study of Maximum Generalized Pattern Matching Problems

Sebastian Ordyniak*

Alexandru Popa

Faculty of Informatics, Masaryk University, Brno, Czech Republic,
sordyniak@gmail.com, popa@fi.muni.cz

September 9, 2014

Abstract

The generalized function matching (GFM) problem has been intensively studied starting with [Ehrenfeucht and Rozenberg, 1979]. Given a pattern p and a text t , the goal is to find a mapping from the letters of p to non-empty substrings of t , such that applying the mapping to p results in t . Very recently, the problem has been investigated within the framework of parameterized complexity [Fernau, Schmid, and Villanger, 2013].

In this paper we study the parameterized complexity of the optimization variant of GFM (called Max-GFM), which has been introduced in [Amir and Nor, 2007]. Here, one is allowed to replace some of the pattern letters with some special symbols “?”, termed wildcards or don’t cares, which can be mapped to an arbitrary substring of the text. The goal is to minimize the number of wildcards used.

We give a complete classification of the parameterized complexity of Max-GFM and its variants under a wide range of parameterizations, such as, the number of occurrences of a letter in the text, the size of the text alphabet, the number of occurrences of a letter in the pattern, the size of the pattern alphabet, the maximum length of a string matched to any pattern letter, the number of wildcards and the maximum size of a string that a wildcard can be mapped to.

1 Introduction

In the generalized function matching problem one is given a text t and a pattern p and the goal is to decide whether there is a match between p and t , where a single letter of the pattern is allowed to match multiple letters of the text (we say that p GF-matches t). For example, if the text is $t = xyxx$ and the pattern is $p = aba$, then a generalized function match (on short, GF-match) is $a \rightarrow x, b \rightarrow yy$, but if $t = xyxz$ and $p = aba$, then there is no GF-match. If, moreover, the matching is required to be injective, then we term the problem generalized parameterized matching (GPM). In [1], Amir and Nor describe applications of GFM in various areas such as software engineering, image searching, DNA analysis, poetry and music analysis, or author validation. GFM is also related to areas such as (un-)avoidable patterns [12], word equations [13] and the ambiguity of morphisms [11].

GFM has a long history starting from 1979. Ehrenfeucht and Rozenberg [7] show that GFM is NP-complete. Independently, Angluin [2, 3] studies a more general variant of GFM where the pattern may contain also letters of the text alphabet. Angluin’s paper received a lot of attention, especially in the learning theory community [16, 17, 19] (see [14] for a survey) but also in many other areas.

Recently, a systematic study of the classical complexity of a number of variants of GFM and GPM under various restrictions has been carried out [8, 18]. It was shown that GFM and GPM remain

*Research funded by Employment of Newly Graduated Doctors of Science for Scientific Excellence (CZ.1.07/2.3.00/30.0009).

NP-complete for many natural restrictions. Moreover, the study of GFM and its variants within the framework of parameterized complexity has recently been initiated [9].

In this paper we study the parameterized complexity of the optimization variant of GFM (called Max-GFM) and its variants, where one is allowed to replace some of the pattern letters with some special symbols “?”, termed wildcards or don’t cares, which can be mapped to an arbitrary substring of the text. The goal is to minimize the number of wildcards used. The problem was first introduced to the pattern matching community by Amir and Nor [1]. They show that if the pattern alphabet has constant size, then a polynomial algorithm can be found, but that the problem is **NP**-complete otherwise. Then, in [4], it is shown the **NP**-hardness of the GFM (without wildcards) and the **NP**-hardness of the GFM when the function f is required to be an injection (named GPM). More specifically, GFM is **NP**-hard even if the text alphabet is binary and each letter of the pattern is allowed to map to at most two letters of the text [4]. In the same paper it is given a \sqrt{OPT} approximation algorithm for the optimization variant of GFM where the goal is to search for a pattern p' that GF-matches t and has the smallest Hamming distance to p . In [5] the optimization versions of GFM and GPM are proved to be **APX**-hard.

OUR RESULTS Before we discuss our results, we give formal definitions of the problems. In the following let t be a text over an alphabet Σ_t and let $p = p_1 \dots p_m$ be a pattern over an alphabet Σ_p . We say that p *GF-matches* t if there is a function $f : \Sigma_p \rightarrow \Sigma_t^+$ such that $f(p_1) \dots f(p_m) = t$. To improve the presentation we will sometimes abuse notation by writing $f(p)$ instead of $f(p_1) \dots f(p_m)$. Let k be a natural number. We say that a pattern p *k-GF-matches* t if there is a text p' over alphabet $\Sigma_p \cup \{?, \dots, ?_k\}$ of Hamming distance at most k from p such that p' GF-matches t .

Problem 1 (MAXIMUM GENERALIZED FUNCTION MATCHING). *Given a text t , a pattern p , and an integer k , decide whether p *k-GF-matches* t .*

The Max-GFM can be seen as the optimization variant of GFM in which we want to replace some of the pattern letters with special wildcard symbols, i.e., the symbols $?, \dots, ?_k$, which can be mapped to any non-empty substring of the text.

We also study the Max-GPM problem. The only difference between Max-GPM and Max-GFM is that for Max-GPM the function f is required to be injective. The notions of GP-matching and k -GP-matching are defined in the natural way, e.g., we say a pattern p *GP-matches* a text t if p *GF-matches* t using an injective function.

In this paper we study the parameterized complexity of the two problems using a wide range of parameters: maximum number of occurrences of a letter in the text $\#\Sigma_t$, maximum number of occurrences of a letter in the pattern $\#\Sigma_p$, size of the text alphabet $|\Sigma_t|$, size of the pattern alphabet $|\Sigma_p|$, the maximum length of a substring of the text that a letter of the pattern alphabet can be mapped to (i.e., $\max_i |f(p_i)|$), the number of wildcard letters $\#?$, and the maximum length of a substring of the text that a wildcard can be mapped to, denoted by $\max |f(?)|$.

Our results are summarized in Table 1. We verified the completeness of our results using a simple computer program. In particular, the program checks for every of the 128 possible combinations of parameters \mathcal{C} that the table contains either: i) a superset of \mathcal{C} under which Max-GFM/GPM is hard (and thus, Max-GFM/GPM is hard if parameterized by \mathcal{C}); or ii) a subset of \mathcal{C} for which Max-GFM/GPM is fpt (and then we have an fpt result for the set of parameters \mathcal{C}). Since some of our results do not hold for both Max-GFM and Max-GPM, we carried out two separate checks, one for Max-GFM and one for Max-GPM.

The paper is organized as follows. In Section 2 we give preliminaries, in Section 3 we present our fixed-parameter algorithms and in Section 4 we show our hardness results.

2 Preliminaries

We define the basic notions of Parameterized Complexity and refer to other sources [6, 10] for an in-depth treatment. A *parameterized problem* is a set of pairs $\langle \mathbb{I}, k \rangle$, the *instances*, where \mathbb{I} is the main part and k the *parameter*. The parameter is usually a non-negative integer. A parameterized problem

$\#\Sigma_t$	$ \Sigma_t $	$\#\Sigma_p$	$ \Sigma_p $	$\max_i f(p_i) $	$\#?$	$\max f(?) $	Complexity
par	par	—	—	—	—	—	FPT (Cor. 3)
—	par	—	par	par	—	—	FPT (Th. 1)
—	par	—	—	par	—	—	FPT only GPM (Cor. 1)
—	—	par	par	par	—	par	FPT (Cor. 2)
—	—	—	par	par	par	par	FPT (Th. 2)
par	—	par	par	par	par	—	W [1]-h (Th. 4)
par	—	par	par	—	par	par	W [1]-h (Th. 7)
par	—	par	—	par	par	par	W [1]-h (Th. 5)
—	par	par	par	—	par	par	W [1]-h ([9, Th. 2.])
—	—	par	par	par	par	—	W [1]-h (Th. 6)
—	—	—	par	par	—	par	W [1]-h (Th. 3)
—	par	par	—	par	par	par	para-NP -h ([1, Cor. 1]),
—	par	par	—	par	—	—	para-NP -h only GFM [8]
—	—	par	—	par	—	—	para-NP -h only GPM [8]

Table 1: Parameterized Complexity of Max-GFM and Max-GPM .

is *fixed-parameter tractable* (*fpt*) if there exists an algorithm that solves any instance $\langle \mathbb{I}, k \rangle$ of size n in time $f(k)n^c$ where f is an arbitrary computable function and c is a constant independent of both n and k . **FPT** is the class of all fixed-parameter tractable decision problems. Because we focus on fixed-parameter tractability of a problem we will sometimes use the notation O^* to suppress exact polynomial dependencies, i.e., a problem with input size n and parameter k can be solved in time $O^*(f(k))$ if it can be solved in time $O(f(k)n^c)$ for some constant c .

Parameterized complexity offers a completeness theory, similar to the theory of NP-completeness, that allows the accumulation of strong theoretical evidence that some parameterized problems are not fixed-parameter tractable. This theory is based on a hierarchy of complexity classes $\mathbf{FPT} \subseteq \mathbf{W}[1] \subseteq \mathbf{W}[2] \subseteq \mathbf{W}[3] \subseteq \dots$ where all inclusions are believed to be strict. An *fpt-reduction* from a parameterized problem P to a parameterized problem Q is a mapping R from instances of P to instances of Q such that (i) $\langle \mathbb{I}, k \rangle$ is a YES-instance of P if and only if $\langle \mathbb{I}', k' \rangle = R(\mathbb{I}, k)$ is a YES-instance of Q , (ii) there is a computable function g such that $k' \leq g(k)$, and (iii) there is a computable function f and a constant c such that R can be computed in time $O(f(k) \cdot n^c)$, where n denotes the size of $\langle \mathbb{I}, k \rangle$.

For our hardness results we will often reduce from the following problem, which is well-known to be **W**[1]-complete [15].

MULTICOLORED CLIQUE

Instance: A k -partite graph $G = \langle V, E \rangle$ with a partition V_1, \dots, V_k of V .

Parameter: The integer k .

Question: Are there nodes v_1, \dots, v_k such that $v_i \in V_i$ and $\{v_i, v_j\} \in E$ for all i and j with $1 \leq i < j \leq k$ (i.e. the subgraph of G induced by $\{v_1, \dots, v_k\}$ is a clique of size k)?

For our hardness proofs we will often make the additional assumptions that (1) $|V_i| = |V_j|$ for every i and j with $1 \leq i < j \leq k$ and (2) $|E_{i,j}| = |E_{r,s}|$ for every i, j, r , and s with $1 \leq i < j \leq k$ and $1 \leq r < s \leq k$, where $E_{i,j} = \{ \{u, v\} \in E \mid u \in V_i \text{ and } v \in V_j \}$ for every i and j as before. To see that MULTICOLORED CLIQUE remains **W**[1]-hard under these additional restrictions we can reduce from MULTICOLORED CLIQUE to its more restricted version using a simple padding construction as follows. Given an instance $\langle G, k \rangle$ of MULTICOLORED CLIQUE we construct an instance of its more restricted version by adding edges (whose endpoints are new vertices) between parts (i.e. V_1, \dots, V_k) that do not already have the maximum number of edges between them and then adding isolated vertices to parts that do not already have the maximum number of vertices.

Even stronger evidence that a parameterized problem is not fixed-parameter tractable can be obtained by showing that the problem remains **NP**-complete even if the parameter is a constant. The class of these problems is called **para-NP**.

A *square* is a string consisting of two copies of the same (non-empty) string. We say that a string is *square-free* if it does not contain a square as a substring.

3 Fixed-parameter Tractable Variants

In this section we show our fixed-parameter tractability results for Max-GFM and Max-GPM. In particular, we show that Max-GFM and Max-GPM are fixed-parameter tractable parameterized by $|\Sigma_t|$, $|\Sigma_p|$, and $\max_i |f(p_i)|$, and also parameterized by $\#?$, $\max |f(?)|$, $|\Sigma_p|$, and $\max_i |f(p_i)|$. We start by showing fixed-parameter tractability for the parameters $|\Sigma_t|$, $|\Sigma_p|$, and $\max_i |f(p_i)|$. We need the following lemma.

Lemma 1. *Given a pattern $p = p_1 \dots p_m$ over an alphabet Σ_p , a text $t = t_1 \dots t_n$ over an alphabet Σ_t , a natural number q , and a function $f : \Sigma_p \rightarrow \Sigma_t^+$, then there is a polynomial time algorithm deciding whether p q -GF/GP-matches t using the function f .*

Proof. If we are asked whether p q -GP-matches t and f is not injective, then we obviously provide a negative answer. Otherwise, we use a dynamic programming algorithm that is similar in spirit to an algorithm in [4]. Let $\Sigma_p = \{a_1, \dots, a_k\}$. For every $0 \leq i \leq j \leq n$, we define the function $g(i, j)$ to be the Hamming GFM/GPM-similarity (i.e., m minus the minimum number of wildcards needed) between $t_1 t_2 \dots t_j$ and $p_1 p_2 \dots p_i$. Then, we obtain the Hamming GFM/GPM-similarity between p and t as $g(m, n)$. Consequently, if $m - g(m, n) > q$, we return NO, otherwise we return YES.

We now show how to recursively compute $g(i, j)$. If $i = 0$, we set $g(i, j) = 0$ and if $i \leq j$, we set:

$$g(i, j) = \max_{1 \leq k \leq j} \{g(i-1, j-k) + I(t_{j-k+1} \dots t_j, f(p_i))\}$$

where $I(s_1, s_2)$ is 1 if the strings s_1 , and s_2 are the same, and 0 otherwise.

We must first show that the dynamic programming procedure computes the right function and then that it runs in polynomial time. We can see immediately that $g(0, i) = 0$ for all i because in this case the pattern is empty. The recursion step of $g(i, j)$ has two cases: If $t_{j-|f(p_i)|+1} \dots t_j = f(p_i)$, then it is possible to map p_i to $f(p_i)$, and we can increase the number of mapped letters by one. Otherwise, we cannot increase the Hamming GFM/GPM-similarity. However, we know that p_i has to be set to a wildcard and therefore we find the maximum of the previous results for different length substrings that the wildcard maps to.

It is straightforward to check that $g(i, j)$ can be computed in cubic time. \square

Theorem 1. *Max-GFM and Max-GPM parameterized by $|\Sigma_t|$, $|\Sigma_p|$, and $\max_i |f(p_i)|$ are fixed-parameter tractable.*

Proof. Let p , t , and q be an instance of Max-GFM or Max-GPM, respectively. The pattern p q -GF/GP-matches t if and only if there is a function $f : \Sigma_p \rightarrow \Sigma_t^+$ such that p q -GF/GP-matches t using f . Hence, to solve Max-GFM/Max-GPM, it is sufficient to apply the algorithm from Lemma 1 to every function $f : \Sigma_p \rightarrow \Sigma_t^+$ that could possibly constitute to a q -GF/GP-matching from p to t . Because there are at most $(|\Sigma_t|)^{\max_i |f(p_i)| |\Sigma_p|}$ such functions f and the algorithm from Lemma 1 runs in polynomial time, the running time of this algorithm is $O^*((|\Sigma_t|)^{\max_i |f(p_i)| |\Sigma_p|})$, and hence fixed-parameter tractable in $|\Sigma_t|$, $|\Sigma_p|$, and $\max_i |f(p_i)|$. \square

Because in the case of Max-GPM it holds that if $|\Sigma_t|$ and $\max_i |f(p_i)|$ is bounded then also Σ_p is bounded by $|\Sigma_t|^{\max_i |f(p_i)|}$, we obtain the following corollary.

Corollary 1. *Max-GPM parameterized by $|\Sigma_t|$ and $\max_i |f(p_i)|$ is fixed-parameter tractable.*

We continue by showing our second tractability result for the parameters $|\Sigma_p|$, $\max_i |f(p_i)|$, $\#?$, and $\max |f(?)|$.

Theorem 2. *Max-GFM and Max-GPM parameterized by $|\Sigma_p|$, $\max_i |f(p_i)|$, $\#?$, $\max |f(?)|$, are fixed-parameter tractable.*

Proof. Let p , t , and q be an instance of Max-GFM or Max-GPM, respectively.

Observe that if we could go over all possible functions $f : \Sigma_p \rightarrow \Sigma_t^+$ that could possibly constitute a q -GF/GP-matching from p to t , then we could again apply Lemma 1 as we did in the proof of Theorem 1. Unfortunately, because $|\Sigma_t|$ is not a parameter, the number of these functions cannot be bounded as easily any more. However, as we will show next it is still possible to bound the number of possible functions solely in terms of the parameters. In particular, we will show that the number of possible substrings of t that any letter of the pattern alphabet can be mapped to is bounded by a function of the parameters. Because also $|\Sigma_p|$ is a parameter this immediately implies a bound (only in terms of the given parameters) on the total number of these functions.

Let $c \in \Sigma_p$ and consider any q -GF/GP-matching from p to t , i.e., a text $p' = p'_1 \dots p'_m$ of Hamming distance at most q to p and a function $f : \Sigma_p \cup \{?, \dots, ?_q\} \rightarrow \Sigma_t^+$ such that $f(p'_1) \dots f(p'_m) = t$. Then either c does not occur in p' or c occurs in p' . In the first case we can assign to c any non-empty substring over the alphabet Σ_t (in the case of Max-GPM one additionally has to ensure that the non-empty substrings over Σ_t that one chooses for distinct letters in Σ_p are distinct). In the second case let p'_i for some i with $1 \leq i \leq m$ be the first occurrence of c in p' , let $\bar{p}'_{i-1} = p'_1 \dots p'_{i-1}$, and let $\bar{p}_{i-1} = p_1 \dots p_{i-1}$. Furthermore, for every $b \in \Sigma_p \cup \{?, \dots, ?_q\}$ and $w \in (\Sigma_p \cup \{?, \dots, ?_q\})^*$, we denote by $\#(b, w)$ the number of times b occurs in w . Then $f(c) = t_{c_s+1} \dots t_{c_s+|f(c)|}$ where $c_s = \sum_{j=1}^{i-1} |f(p'_j)|$, which implies that the value of $f(c)$ is fully determined by c_s and $|f(c)|$. Because the number of possible values for $|f(c)|$ is trivially bounded by the parameters (it is bounded by $\max_i |f(p_i)|$), it remains to show that also c_s is bounded by the given parameters.

Because $c_s = \sum_{j=1}^{i-1} |f(p'_j)| = (\sum_{b \in \Sigma_p \cup \{?, \dots, ?_q\}} \#(b, \bar{p}'_{i-1}) |f(b)|)$, we obtain that the value of c_s is fully determined by the values of $\#(b, \bar{p}'_{i-1})$ and $|f(b)|$ for every $b \in \Sigma_p \cup \{?, \dots, ?_q\}$. For every $? \in \{?, \dots, ?_q\}$ there are at most 2 possible values for $\#(?, \bar{p}'_{i-1})$ (namely 0 and 1) and there are at most $\max |f(?)|$ possible values for $|f(?)|$. Similarly, for every $b \in \Sigma_p$ there are at most $q + 1$ possible values for $\#(b, \bar{p}'_{i-1})$ (the values $\#(b, \bar{p}_{i-1}) - q, \dots, \#(b, \bar{p}_{i-1})$) and there are at most $\max_i |f(p_i)|$ possible values for $|f(b)|$. Hence, the number of possible values for c_s is bounded in terms of the parameters, as required. \square

Since $|\Sigma_p|$ and $\#\Sigma_p$ together bound $\#?$, we obtain the following corollary.

Corollary 2. *Max-GFM and Max-GPM parameterized by $\#\Sigma_p$, $|\Sigma_p|$, $\max_i |f(p_i)|$, and $\max |f(?)|$ are fixed-parameter tractable.*

Furthermore, because all considered parameters can be bounded in terms of the parameters $\#\Sigma_t$ and $|\Sigma_t|$, we obtain the following corollary as a consequence of any of our above fpt-results.

Corollary 3. *Max-GFM and Max-GPM parameterized by $\#\Sigma_t$ and $|\Sigma_t|$ are fixed-parameter tractable.*

4 Hardness Results

In this subsection we give our hardness results for Max-GFM and Max-GPM.

Theorem 3. *Max-GFM and Max-GPM are $\mathbf{W}[1]$ -hard parameterized by $|\Sigma_p|$, $\max_i |f(p_i)|$, and $\max |f(?)|$ (even if $\max_i |f(p_i)| = 1$ and $\max |f(?)| = 2$).*

We will show the theorem by a parameterized reduction from MULTICOLORED CLIQUE. To simplify the proof we will reduce to the variant of Max-GFM and Max-GPM, where we are allowed to map wildcards to the empty string. It is however straightforward to adapt the proof to the original versions of Max-GFM and Max-GPM. Hence, in the following, whenever we refer to Max-GFM and Max-GPM, we mean the version of Max-GFM and Max-GPM, where wildcards can be mapped to the empty string.

Let $G = (V, E)$ be a k -partite graph with partition V_1, \dots, V_k of V . Let $E_{i,j} = \{ \{u, v\} \in E \mid u \in V_i \text{ and } v \in V_j \}$ for every i and j with $1 \leq i < j \leq k$. Again, as we stated in the preliminaries we can assume that $|V_i| = n$ and $|E_{i,j}| = m$ for every i and j with $1 \leq i < j \leq k$.

Let $V_i = \{v_1^i, \dots, v_n^i\}$ and $E_{i,j} = \{e_1^{i,j}, \dots, e_m^{i,j}\}$. We construct a text t and a pattern p from G and k such that p r -GF/GP-matches t with $r = \binom{k}{2}(8(m-1))$ if and only if G has a k -clique. We set $\Sigma_t = \{;, -, \#, \square\} \cup \{v_i^j \mid 1 \leq i \leq n \text{ and } 1 \leq j \leq k\}$ and $\Sigma_p = \{;, -, \#, \square\} \cup \{V_i \mid 1 \leq i \leq k\}$.

For an edge $e \in E$ between v_l^i and v_k^j where $1 \leq i < j \leq k$ and $1 \leq l, k \leq n$, we write $\mathbf{vt}(e)$ to denote the text $v_l^i - v_k^j$. For $l \in \Sigma_p \cup \Sigma_t$ and $i \in \mathbb{N}$ we write $\mathbf{rp}(l, i)$ to denote the text consisting of repeating the letter l exactly i times. We first define a preliminary text t' as follows.

$$\begin{aligned} & \#; \mathbf{vt}(e_1^{1,2}); \dots; \mathbf{vt}(e_m^{1,2}); \# \dots \#; \mathbf{vt}(e_1^{1,k}); \dots; \mathbf{vt}(e_m^{1,k}); \\ & \#; \mathbf{vt}(e_1^{2,3}); \dots; \mathbf{vt}(e_m^{2,3}); \# \dots \#; \mathbf{vt}(e_1^{2,k}); \dots; \mathbf{vt}(e_m^{2,k}); \\ & \dots \\ & \#; \mathbf{vt}(e_1^{k-1,k}); \dots; \mathbf{vt}(e_m^{k-1,k}); \# \end{aligned}$$

We also need to define a preliminary pattern p' as follows.

$$\begin{aligned} & \# \mathbf{rp}(\square, 4(m-1)); V_1 - V_2; \mathbf{rp}(\square, 4(m-1)) \# \dots \\ & \# \mathbf{rp}(\square, 4(m-1)); V_1 - V_k; \mathbf{rp}(\square, 4(m-1)) \\ & \# \mathbf{rp}(\square, 4(m-1)); V_2 - V_3; \mathbf{rp}(\square, 4(m-1)) \# \dots \\ & \# \mathbf{rp}(\square, 4(m-1)); V_2 - V_k; \mathbf{rp}(\square, 4(m-1)) \\ & \dots \\ & \# \mathbf{rp}(\square, 4(m-1)); V_{k-1} - V_k; \mathbf{rp}(\square, 4(m-1)) \# \end{aligned}$$

We obtain t from t' and p from p' by preceding t' and p' , respectively, with the following text or pattern, respectively.

$$\mathbf{rp}(\square, 2r+1) \mathbf{rp}(;, 2r+1) \mathbf{rp}(-, 2r+1) \mathbf{rp}(\#, 2r+1)$$

This completes the construction of t and p . Clearly, t and p can be constructed from G and k in fpt-time (even polynomial time). Furthermore, $|\Sigma_p| = k+4$, as required. It remains to show that G has a k -clique if and only if p r -GF/GP-matches t .

Lemma 2. *If G has a k -clique then p r -GF/GP-matches t .*

Proof. Let $\{v_{h_1}^1, \dots, v_{h_k}^k\}$ be the vertices and $\{e_{h_i,j}^{i,j} \mid 1 \leq i < j \leq k\}$ be the edges of a k -clique of G with $1 \leq h_j \leq n$ and $1 \leq h_{i,j} \leq m$ for every i and j with $1 \leq i < j \leq k$.

The function f that r -GF/GP-matching p to t is defined as follows: $f(\square) = \square$, $f(;) = ;$, $f(-) = -$, $f(\#) = \#$, $f(V_i) = v_{h_1}^i$, for every i with $1 \leq i \leq k$.

We put r wildcards on the last r occurrences of \square in p , i.e., every occurrence of \square that corresponds to an occurrence in p' . Then length of the text the wildcards are mapped to is determined as follows. For an edge $e_{h_i,j}^{i,j}$ look at the “block” in p that corresponds to the edge, i.e., the block:

$$\# \mathbf{rp}(\square, 4(m-1)); V_i - V_j; \mathbf{rp}(\square, 4(m-1))$$

The first $4(m - h_{i,j})$ occurrences of \square (in this block) are replaced with a wildcard which is mapped to a text of length 0, the last $4(m - h_{i,j})$ occurrences of \square are replaced with a wildcard which is mapped to a text of length 2, and all other occurrences of \square are replaced with a wildcard that is mapped to a text of length 1. It is straightforward to check that f together with the mapping of the wildcards maps the pattern p to the text t . \square

For the reverse direction we need the following intermediate claims.

Claim 1. *For any function f that r -GF/GP-matches p to t it holds that: $f(\square) = \square$, $f(;) = ;$, $f(-) = -$, and $f(\#) = \#$.*

Proof. We show that $f(\square) = \square$ since the remaining cases are similar. Because the pattern p starts with $2r+1$ repetitions of the letter \square , it follows that at least 1 of these occurrences of \square is not replaced with a wildcard. Because every letter of p is replaced by at most 2 letters of the text the first occurrence of \square that is not replaced by a wildcard is mapped to a letter of the text at position at most $2r$, i.e., a \square . This concludes the proof of the claim. \square

Claim 2. *Any r -GF/GP-matching of p to t replaces exactly the last r occurrences of \square in p with wildcards.*

Proof. It follows from the previous claim that $f(\square) = \square$ for any function that r -GF/GP-matches p to t . Because every letter of p that is not replaced with a wildcard is replaced with exactly 1 letter from the text, it follows that the first occurrence of \square in p that corresponds to an occurrence of \square in p' is mapped to (if it is not replaced with a wildcard) to a letter of the text at position at least $7r+5$. However, since the text t does not contain the letter \square after position $2r+1$, this occurrence of \square in p (and all other occurrences of \square in p that follow) has to be replaced with a wildcard. Since p' contains exactly r occurrences of \square the only letters of p that are replaced with wildcards are these occurrences of \square . \square

Lemma 3. *If p r -GF/GP-matches t then G has a k -clique.*

Proof. Let f be a function that r -GF/GP-matches p to t . Because of Claim 9, it holds that $f(\square) = \square$, $f(;) = ;$, $f(-) = -$, and $f(\#) = \#$. Furthermore, because of Claim 9 the only letters in p that are replaced with wildcards are the last r occurrences of \square in p . Because the number of occurrences of the letter $\#$ is the same in t and p each occurrence of $\#$ in p has to be mapped to its corresponding occurrence in t . It follows that for every i and j with $1 \leq i < j \leq k$ the “block”

$$\mathbf{rp}(\square, 4(m-1)); V_i - V_j; \mathbf{rp}(\square, 4(m-1))$$

in p has to be mapped to the corresponding “block”

$$; \mathbf{vt}(e_1^{i,j}); \dots; \mathbf{vt}(e_m^{i,j});$$

in t . Hence, the part $V_i - V_j$ has to be mapped to $\mathbf{vt}(e_l^{i,j})$ for every $1 \leq l \leq m$. Consequently, the set $\{f(V_i) \mid 1 \leq i \leq k\}$ is a k -clique of G . \square

This concludes the proof of Theorem 3.

Theorem 4. *Max-GFM and Max-GPM are $\mathbf{W}[1]$ -hard parameterized by $\#\Sigma_t$, $\#\Sigma_p$, $|\Sigma_p|$, $\max_i |f(p_i)|$, and $\#?$.*

We will show the theorem by a parameterized reduction from MULTICOLORED CLIQUE. Let $G = (V, E)$ be a k -partite graph with partition V_1, \dots, V_k of V . Let $E_{i,j} = \{\{u, v\} \in E \mid u \in V_i \text{ and } v \in V_j\}$ for every i and j with $1 \leq i < j \leq k$. Again, as we stated in the preliminaries we can assume that $|V_i| = n$ and $|E_{i,j}| = m$ for every i and j with $1 \leq i < j \leq k$.

Let $V_i = \{v_1^i, \dots, v_n^i\}$ and $E_{i,j} = \{e_1^{i,j}, \dots, e_m^{i,j}\}$, and $k' = 2\binom{k}{2} + k(k+2)$. We construct a text t over alphabet Σ_t and a pattern p over alphabet Σ_p from G and k such that p k' -GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 1$ if and only if G has a k -clique. The alphabet Σ_t consists of:

- the letter $\#$ (used as a separator);
- the letter $+$ (used to forced the wildcards);
- one letter a_e for every $e \in E$ (representing the edges of G);
- one letter $\#_i$ for every i with $1 \leq i \leq n$ (used as special separators that group edges from the same vertex);
- the letters $l_{i,j}$, $r_{i,j}$, l_i , r_i for every i and j with $1 \leq i < j \leq k$ (used as dummy letters to ensure injectivity for GPM);

- the letter d_e^v and d^v for every $e \in E$ and $v \in V(G)$ with $v \in e$ (used as dummy letters to ensure injectivity for GPM).

We set $\Sigma_p = \{\#, D\} \cup \{E_{i,j} \mid 1 \leq i < j \leq k\}$.

For a vertex $v \in V$ and j with $1 \leq j \leq k$ we denote by $E_j(v)$ the set of edges of G that are incident to v and whose other endpoint is in V_j . Furthermore, for a vertex $v \in V(G)$, we write $\mathbf{e}(v)$ to denote the text $\mathbf{el}(v, E_1(v)) \cdots \mathbf{el}(v, E_k(v))d^v$, where $\mathbf{el}(v, E')$, for vertex v and a set E' of edges with $E' = \{e_1, \dots, e_l\}$, is the text $d_{e_1}^v e_{e_1} d_{e_2}^v e_{e_2} \cdots d_{e_l}^v a_{e_l}$.

We first define the following preliminary text and pattern strings. Let t_1 be the text:

$$\begin{aligned} & \#l_{1,2}a_{e_1^{1,2}} \cdots a_{e_m^{1,2}}r_{1,2}\# \cdots \#l_{1,k}a_{e_1^{1,k}} \cdots a_{e_m^{1,k}}r_{1,k} \\ & \#l_{2,3}a_{e_1^{2,3}} \cdots a_{e_m^{2,3}}r_{2,3}\# \cdots \#l_{2,k}a_{e_1^{2,k}} \cdots a_{e_m^{2,k}}r_{2,k} \\ & \quad \dots \\ & \#l_{k-1,k}a_{e_1^{k-1,k}} \cdots a_{e_m^{k-1,k}}r_{k-1,k} \end{aligned}$$

Let t_2 be the text:

$$\begin{aligned} & \#l_1\#_1\mathbf{e}(v_1^1)\#_1 \cdots \#_n\mathbf{e}(v_n^1)\#_nr_1 \\ & \quad \dots \\ & \#l_k\#_1\mathbf{e}(v_1^k)\#_1 \cdots \#_n\mathbf{e}(v_n^k)\#_nr_k\# \end{aligned}$$

Let p_1 be the pattern:

$$\begin{aligned} & \#DE_{1,2}D\# \dots \#DE_{1,k}D \\ & \#DE_{2,3}D\# \dots \#DE_{2,k}D \\ & \quad \dots \\ & \#DE_{k-1,k}D \end{aligned}$$

For i, j with $1 \leq i, j \leq k$, let $I(i, j)$ be the letter $E_{i,j}$ if $i < j$, the letter $E_{j,i}$ if $i > j$ and the empty string if $i = j$. We define $\mathbf{p}(1)$ to be the pattern:

$$A_1DI(1,2)DI(1,3) \cdots DI(1,k)DA_1$$

, we define $\mathbf{p}(k)$ to be the pattern:

$$A_kDI(k,1)DI(k,2) \cdots DI(k,k-1)DA_k$$

, and for every i with $1 < i < k$, we define $\mathbf{p}(i)$ to be the pattern:

$$A_iDI(i,1)DI(i,2) \cdots DI(i,i-1)DI(i,i+1) \cdots DI(i,k)DA_i$$

Then p_2 is the pattern:

$$\#L_1\mathbf{p}(1)R_1\# \cdots \#L_k\mathbf{p}(k)R_k\#$$

Let $r = 2(k' + 1)$. For $l \in \Sigma_p \cup \Sigma_t$ and $i \in \mathbb{N}$ we write $\mathbf{rp}(l, i)$ to denote the text consisting of repeating the letter l exactly i times. We also define t_0 to be the text $\#\mathbf{rp}(+, r)$ and p_0 to be the pattern $\#\mathbf{rp}(D, r)$. Then, t is the concatenation of t_0 , t_1 and t_2 and p is a concatenation of p_0 , p_1 and p_2 .

This completes the construction of t and p . Clearly, t and p can be constructed from G and k in fpt-time (even polynomial time). Furthermore, $\#\Sigma_t = r$, $\#\Sigma_p = r + k'$, $|\Sigma_p| = \binom{k}{2} + k + 2$ and hence bounded by k , as required. It remains to show that G has a k -clique if and only if p k' -GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 1$.

Lemma 4. *If G has a k -clique then p k' -GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 1$.*

Proof. Let $\{v_{h_1}^1, \dots, v_{h_k}^k\}$ be the vertices and $\{e_{h_{i,j}}^{i,j} \mid 1 \leq i < j \leq k\}$ be the edges of a k -clique of G with $1 \leq h_i \leq n$ and $1 \leq h_{i,j} \leq m$ for every i and j with $1 \leq i < j \leq k$.

We put k' wildcards on the last k' occurrences of D in p . The mapping of these wildcards is defined very similar to the mapping of the letters $L_{i,j}$, $R_{i,j}$, L_i , R_i , and $D_{i,j}$ in the proof of Lemma 12 and will not be repeated here. Using this mapping ensures that every wildcard is mapped to a non-empty substring of t and no two wildcards are mapped to the same substring of t .

We define the function f that k' -GF/GP-matches p to t as follows: We set $f(\#) = \#$ and $f(D) = +$. Moreover, for every i and j with $1 \leq i < j \leq k$, we set $f(E_{i,j}) = a_{e_{h_{i,j}}^{i,j}}$ and $f(A_i) = \#_i$.

It is straightforward to check that f together with above mapping for the wildcards k' -GF/GP-matches p to t . \square

Claim 3. *Let f be a function that k' -GF/GP-matches p to t with $\max_{p \in \Sigma_p} |f(p)| = 1$, then: $f(\#) = \#$ and $f(D) = +$. Moreover, all wildcards have to be placed on all the k' occurrences of D in p' .*

Proof. We first show that $f(D) = +$. Observe that the only squares in the string t are contained in t_0 (recall the definition of square-free from Section 2). It follows that every two consecutive occurrences of pattern letters in p_0 have to be mapped to a substring of t_0 . Because there are $2(k' + 1)$ occurrences of D in p_0 it follows that at least two consecutive occurrences of D in p_0 are not replaced with wildcards and hence D has to be mapped to a substring of t_0 . Furthermore, since all occurrences of D are at the end of p_0 , we obtain that D has to be mapped to $+$, as required. Because all occurrences of D in p' have to be mapped to substrings of the concatenation of t_1 and t_2 , but these strings do not contain the letter $+$, it follows that all the k' occurrences of D in p_1 and p_2 have to be replaced by wildcards. Since we are only allowed to use at most k' wildcards, this shows the second statement of the claim. Since no wildcards are used to replace letters in p_0 it now easily follows that $f(\#) = \#$. \square

Lemma 5. *If p k' -GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 1$, then G has a k -clique.*

Proof. Let f be a function that k' -GF/GP-matches p to t with $\max_{p \in \Sigma_p} |f(p)| = 1$. Because of Claim 3, we know that $f(\#) = \#$ and that no occurrence of $\#$ in p is replaced by a wildcard. Because t and p have the same number of occurrences of $\#$, it follows that the i -th occurrences of $\#$ in p has to be mapped to the i -th occurrence of $\#$ in t . We obtain that:

- (1) For every i, j with $1 \leq i < j \leq k$, the substring $DE_{i,j}D$ of p has to be mapped to the substring $l_{i,j}a_{e_{h_{i,j}}^{i,j}} \cdots a_{e_{h_{j,i}}^{j,i}}r_{i,j}$ of t .
- (2) For every i with $1 \leq i \leq k$, the substring $L_i \mathbf{p}(i) R_i$ of p has to be mapped to the substring $l_i \#_1 \mathbf{e}(v_1^i) \#_1 \cdots \#_n \mathbf{e}(v_n^i) \#_n r_i$ of t .

Because for every i with $1 \leq i \leq k$ the letters $\#_j$ are the only letters that occur more than once in the substring $l_i \#_1 \mathbf{e}(v_1^i) \#_1 \cdots \#_n \mathbf{e}(v_n^i) \#_n r_i$ of t , we obtain from (2) that A_i has to be mapped to $\#_j$ for some j with $1 \leq j \leq n$. Consequently:

- (3) for every i with $1 \leq i \leq k$, the substring $\mathbf{p}(i)$ of p has to be mapped to a substring $\#_j \mathbf{e}(v_j^i) \#_j$ of t for some j with $1 \leq j \leq n$.

It follows from (1) that for every i, j with $1 \leq i < j \leq k$, $f(E_{i,j})$ is mapped to an edge between V_i and V_j . Furthermore, because of (3) it follows that for every i with $1 \leq i \leq k$, it holds that the edges mapped to any $E_{l,r}$ with $1 \leq l < r \leq k$ such that $l = i$ or $r = i$ have the same endpoint in V_i . Hence, the set of edges mapped to the letters $E_{i,j}$ for $1 \leq i < j \leq k$ form a k -clique of G . \square

This concludes the proof of Theorem 4.

Theorem 5. *(Max-)GFM and (Max-)GPM are $\mathbf{W}[1]$ -hard parameterized by $\#\Sigma_t$, $\#\Sigma_p$, $\max_i |f(p_i)|$, $\#?$, and $\max |f(?)|$.*

We will show the above theorem by a parameterized reduction from MULTICOLORED CLIQUE. Let $G = (V, E)$ be a k -partite graph with partition V_1, \dots, V_k of V . Let $E_{i,j} = \{\{u, v\} \in E \mid u \in V_i \text{ and } v \in V_j\}$ for every i and j with $1 \leq i < j \leq k$. Again, as we stated in the preliminaries we can assume that $|V_i| = n$ and $|E_{i,j}| = m$ for every i and j with $1 \leq i < j \leq k$.

Let $V_i = \{v_1^i, \dots, v_n^i\}$ and $E_{i,j} = \{e_1^{i,j}, \dots, e_m^{i,j}\}$. For a vertex $v \in V_i$ and j with $1 \leq j \leq k$ we denote by $E_j(v)$ the set of edges of G that are incident to v and whose other endpoint is in V_j .

We construct a text t over alphabet Σ_t and a pattern p over alphabet Σ_p from G and k such that the following two conditions hold:

(C1) the parameters $\#\Sigma_t$ and $\#\Sigma_p$ are bounded by k (note the parameters $\#?$ and $\max |f(?)|$ are bounded since we consider GFM and GPM).

(C2) p GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| \leq 2$ if and only if G has a k -clique.

Let $r = 2kn(n-1) + 2n + (k-1)m - 1$. The alphabet Σ_t consists of (1) the letter $\#$, (2) the letters $l_l^{i,j}$ and $r_l^{i,j}$ for every $1 \leq i < j \leq k$ and $1 \leq l \leq m-1$, (3) the letters $l_l^{v,j}$ and $r_l^{v,j}$ for every $v \in V_i$, $1 \leq j \leq k$, and $1 \leq l \leq n-1$, where $1 \leq i \leq k$ and $j \neq i$, (4) the letters l_l^i and r_l^i for every $1 \leq i \leq k$ and $1 \leq l \leq r$, (5) the letter $e_l^{i,j}$ for every $1 \leq i < j \leq k$ and $1 \leq l \leq m$, and (6) the letter $\#_i$ for every $1 \leq i \leq n$.

The alphabet Σ_p consists of (1) the letter $\#$, (2) the letters $L_l^{i,j}$ and $R_l^{i,j}$ for every $1 \leq i < j \leq k$ and $1 \leq l \leq m-1$, (3) the letters $LL_l^{i,j}$ and $RR_l^{i,j}$ for every $1 \leq i, j \leq k$ with $i \neq j$, and $1 \leq l \leq n-1$, (4) the letters L_l^i and R_l^i for every $1 \leq i \leq k$ and $1 \leq l \leq r$, (5) the letter $E_{i,j}$ for every $1 \leq i < j \leq k$, and (6) the letter A_i for every $1 \leq i \leq n$.

For a symbol l and $i \in \mathbb{N}$, we write $\mathbf{enu}(l, i)$ to denote the text $l_1 \dots l_i$.

Furthermore, for a vertex $v \in V(G)$ and i with $1 \leq i \leq k$, we write $\mathbf{e}(v, i)$ to denote the text $\mathbf{el}(E_i(v))$, where $\mathbf{el}(E')$ (for a set of edges E') is a list of all the letters in Σ_t that correspond to the edges in E' .

We first define the following preliminary text and pattern strings. For i and j with $1 \leq i < j \leq k$, we denote by $\mathbf{t}(i, j)$ the text $\mathbf{enu}(l^{i,j}, m-1)\mathbf{enu}(e^{i,j}, m)\mathbf{enu}(r^{i,j}, m-1)$. We define t_1 to be the text:

$$\begin{aligned} & \# \mathbf{t}(1, 2) \# \dots \# \mathbf{t}(1, k) \\ & \# \mathbf{t}(2, 3) \# \dots \# \mathbf{t}(2, k) \\ & \dots \\ & \# \mathbf{t}(k-1, k) \# \end{aligned}$$

For a vertex $v \in V_i$, and j with $1 \leq j \leq k$, we denote by $\mathbf{t}(v, j)$ the text $\mathbf{enu}(l^{v,j}, n-1)\mathbf{e}(v, j)\mathbf{enu}(r^{v,j}, n-1)$ if $j \neq i$ and the empty text if $j = i$. Furthermore, we denote by $\mathbf{t}(v)$ the text $\mathbf{t}(v, 1) \dots \mathbf{t}(v, k)$. Let t_2 be the text:

$$\begin{aligned} & \mathbf{enu}(l^1, r) \#_1 \mathbf{t}(v_1^1) \#_1 \dots \#_n \mathbf{t}(v_n^1) \#_n \mathbf{enu}(r^1, r) \\ & \dots \\ & \# \mathbf{enu}(l^k, r) \#_1 \mathbf{t}(v_1^k) \#_1 \dots \#_n \mathbf{t}(v_n^k) \#_n \mathbf{enu}(r^k, r) \end{aligned}$$

For i and j with $1 \leq i < j \leq k$, we denote by $\mathbf{p}(i, j)$ the pattern $\mathbf{enu}(L^{i,j}, m-1)E_{i,j}\mathbf{enu}(R^{i,j}, m-1)$. Let p_1 be the pattern:

$$\begin{aligned} & \# \mathbf{p}(1, 2) \# \dots \# \mathbf{p}(1, k) \\ & \# \mathbf{p}(2, 3) \# \dots \# \mathbf{p}(2, k) \\ & \dots \\ & \# \mathbf{p}(k-1, k) \# \end{aligned}$$

For i, j with $1 \leq i, j \leq k$, let $I(i, j)$ be the letter $E_{i,j}$ if $i < j$, the letter $E_{j,i}$ if $i > j$ and the empty string if $i = j$. Furthermore, let $\mathbf{pe}(i, j)$ be the pattern $\mathbf{enu}(LL^{i,j}, n-1)I(i, j)\mathbf{enu}(RR^{i,j}, n-1)$ if $i \neq j$ and the empty pattern otherwise. Let p_2 be the pattern:

$$\begin{aligned}
& \text{enu}(L^1, r) A_1 \text{pe}(1, 1) \cdots \text{pe}(1, k) A_1 \text{enu}(R^1, r) \\
& \quad \dots \\
& \# \text{enu}(L^k, r) A_k \text{pe}(k, 1) \cdots \text{pe}(k, k) A_k \text{enu}(R^k, r)
\end{aligned}$$

We also define t_0 to be the text $\#\#$ and p_0 to be the pattern $\#\#$. Then, t is the concatenation of t_0 , t_1 and t_2 and p is a concatenation of p_0 , p_1 and p_2 .

This completes the construction of t and p . Clearly, t and p can be constructed from G and k in fpt-time (even polynomial time). Furthermore, because $\#\Sigma_t = \binom{k}{2} + k + 2$, $|\#\Sigma_p| = \binom{k}{2} + k + 2$, condition (C1) is satisfied. To show the remaining condition (C2) we need the following intermediate lemmas and claims.

Lemma 6. *If G has a k -clique then p GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 2$.*

Proof. Let $\{v_{h_1}^1, \dots, v_{h_k}^k\}$ be the vertices and $\{e_{h_i, j}^{i, j} \mid 1 \leq i < j \leq k\}$ be the edges of a k -clique of G with $1 \leq h_j \leq n$ and $1 \leq h_{i, j} \leq m$ for every i and j with $1 \leq i < j \leq k$.

We first give the GF/GP-matching function f for the letters in Σ_p that occur more than once in p as follows: We set $f(\#) = \#$, $f(E_{i, j}) = e_{h_i, j}^{i, j}$, and $f(A_i) = \#_{h_i}$, for every i and j with $1 \leq i < j \leq k$. Informally, we will map the remaining letters in Σ_p to substrings of t of length between 1 and 2 in such a way that the occurrences of the letters $\#$, $E_{i, j}$, and A_i are placed over the right positions in the text t . More formally, we define f for the remaining letters in Σ_p as follows:

- For every $1 \leq i < j \leq k$, we define $f(L_l^{i, j})$ in such a way that $|f(L_l^{i, j})| = 2$ for every $1 \leq l \leq h_{i, j} - 1$ and $|f(L_l^{i, j})| = 1$ for every $h_{i, j} - 1 < l \leq m - 1$.
- For every $1 \leq i < j \leq k$, we define $f(R_l^{i, j})$ in such a way that $|f(R_l^{i, j})| = 1$ for every $1 \leq l \leq h_{i, j} + 1$ and $|f(R_l^{i, j})| = 2$ for every $h_{i, j} + 1 < l \leq m - 1$.
- For every $1 \leq i, j \leq k$ with $i \neq j$, we define $f(LL_l^{i, j})$ in such a way that $f(LL_l^{i, j}) = 2$ for every $1 \leq l \leq s - 1$, where s is the position of $e_{h_i, j}^{i, j}$ in $t(v_{h_i}, j)$ and $f(LL_l^{i, j}) = 1$ for every $s < l \leq n - 1$.
- For every $1 \leq i, j \leq k$ with $i \neq j$, we define $f(RR_l^{i, j})$ in such a way that $f(RR_l^{i, j}) = 1$ for every $1 \leq l \leq s + 1$, where s is the position of $e_{h_i, j}^{i, j}$ in $t(v_{h_i}, j)$ and $f(RR_l^{i, j}) = 2$ for every $s + 1 < l \leq n - 1$.
- For every $1 \leq i \leq k$, we define $f(L_l^i)$ in such a way that $|f(L_l^i)| = 2$ for every $1 \leq l \leq s - 1$, where s is position of $\#_{h_i}$ in the substring $\#_1 t(v_1^i) \#_1 \cdots \#_n t(v_n^i) \#_n$ of t and $|f(L_l^i)| = 1$ for every $s < l \leq r$.
- For every $1 \leq i \leq k$, we define $f(R_l^i)$ in such a way that $|f(R_l^i)| = 1$ for every $1 \leq l \leq s + 1$, where s is position of $\#_{h_i}$ in the substring $\#_1 t(v_1^i) \#_1 \cdots \#_n t(v_n^i) \#_n$ of t and $|f(R_l^i)| = 2$ for every $s + 1 < l \leq r$.

It is now straightforward to check that f GF/GP-matches p to t and $\max_{p \in \Sigma_p} |f(p)| = 2$, as required. \square

To prove the reverse direction we need the following intermediate claim.

Claim 4. *For any function f that k' -GF/GP-matches p to t it holds that: $f(;) = ;$, $f(\#) = \#$, and $f(Q) = +$. Moreover, all wildcards have to be placed on all the k' occurrences of Q in p .*

Proof. We first show that $f(Q) = +$. Observe that the concatenation of the strings t_1 and t_2 is square-free (recall the definition of square-free from Section 2). It follows that every two consecutive occurrences of pattern letters in p_0 have to be mapped to a substring of t_0 . Because there are $2(k' + 1)$ occurrences of Q in p_0 it follows that at least two consecutive occurrences of Q in p_0 are not replaced with wildcards and hence Q has to be mapped to a substring of t_0 . Furthermore, since all occurrences of Q are at the end of p_0 , we obtain that Q has to be mapped to $+$, as required. Because all occurrences of Q in p_1 and p_2 have to be mapped to substrings of the concatenation of t_1 and t_2 but neither t_1 nor t_2 contain the letter $+$, it follows that all the k' occurrences of Q in p_1 and p_2 have to be replaced by wildcards. Since we are only allowed to use at most k' wildcards, this shows the second statement of the claim. Since no wildcards are used to replace letters in p_0 it now also follows that $f(;) = ;$; and $f(\#) = \#$. \square

Lemma 7. *If p GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 2$, then G has a k -clique.*

Proof. Let f be the function that GF/GP-matches p to t with $\max_{p \in \Sigma_p} |f(p)| = 2$. We first show that $f(\#) = \#$. Suppose for a contradiction that $f(\#) \neq \#$. Because t and p start with $\#\#$ it follows that $f(\#)$ is a string that starts with $\#\#$. However, t does not contain any other occurrence of the string $\#\#$ and hence the remaining occurrences of $\#$ in p cannot be matched by f .

Because t and p have the same number of occurrences of $\#$, it follows that the i -th occurrences of $\#$ in p has to be mapped to the i -th occurrence of $\#$ in t . We obtain that:

- (1) for every i, j with $1 \leq i < j \leq k$, the substring $\mathbf{p}(i, j)$ of p has to be mapped to the substring $\mathbf{t}(i, j)$ of t .
- (2) for every i with $1 \leq i \leq k$, the substring $\mathbf{enu}(L^i, r)A_i\mathbf{pe}(i, 1) \cdots \mathbf{pe}(i, k)A_i\mathbf{enu}(R^i, r)$ of p has to be mapped to the substring $\mathbf{enu}(l^i, r)\#_1\mathbf{t}(v_1^i)\#_1 \cdots \#_n\mathbf{t}(v_n^i)\#_n\mathbf{enu}(r^i, r)$ of t .

Because for every i with $1 \leq i \leq k$ the letters $\#_j$ are the only letters that occur more than once in the substring $\mathbf{enu}(l^i, r)\#_1\mathbf{t}(v_1^i)\#_1 \cdots \#_n\mathbf{t}(v_n^i)\#_n\mathbf{enu}(r^i, r)$ of t , we obtain that A_i has to be mapped to $\#_j$ for some j with $1 \leq j \leq n$. Consequently:

- (3) for every i with $1 \leq i \leq k$, the substring $A_i\mathbf{pe}(i, 1) \cdots \mathbf{pe}(i, k)A_i$ of p has to be mapped a substring $\#_j\mathbf{t}(v_j^i)\#_j$ of t for some j with $1 \leq j \leq n$.

It follows from (1) that for every i, j with $1 \leq i < j \leq k$, the function f maps $E_{i,j}$ to edges between V_i and V_j . W.l.o.g. we can assume that $E_{i,j}$ is mapped to exactly one such edge because mapping it to many edges makes it only harder to map the following occurrences of $E_{i,j}$ in p . Because of (3) it follows that for every i with $1 \leq i \leq k$, it holds that the edges mapped to any $E_{l,r}$ with $1 \leq l < r \leq k$ such that $l = i$ or $r = i$ have the same endpoint in V_i . Hence, the set of edges mapped to all the letters $E_{i,j}$ for $1 \leq i < j \leq k$ form a k -clique of G . \square

This concludes the proof of Theorem 5.

Theorem 6. *Max-GFM and Max-GPM are $\mathbf{W}[1]$ -hard parameterized by $\#\Sigma_p$, $|\Sigma_p|$, $\max_i |f(p_i)|$, and $\#?$ (even if $\max_i |f(p_i)| = 1$).*

We will show the above theorem by a parameterized reduction from MULTICOLORED CLIQUE. Let $G = (V, E)$ be a k -partite graph with partition V_1, \dots, V_k of V . Let $E_{i,j} = \{\{u, v\} \in E \mid u \in V_i \text{ and } v \in V_j\}$ for every i and j with $1 \leq i < j \leq k$. As we stated in the preliminaries we can assume that $|V_i| = n$ and $|E_{i,j}| = m$ for every i and j with $1 \leq i < j \leq k$.

Let $V_i = \{v_1^i, \dots, v_n^i\}$, $E_{i,j} = \{e_1^{i,j}, \dots, e_m^{i,j}\}$, and $k' = 2\binom{k}{2}$. We construct a text t over alphabet Σ_t and a pattern p over alphabet Σ_p from G and k in polynomial time such that:

- (C1) the parameters $\#\Sigma_p$, $|\Sigma_p|$, and $\#?$ can be bounded as a function of k .
- (C2) p k' -GF/GP-matches t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 1$ if and only if G has a k -clique.

We set $\Sigma_t = \{;, -, \#, +\} \cup \{l_{i,j}, r_{i,j} \mid 1 \leq i < j \leq k\} \cup \{v_i^j \mid 1 \leq i \leq n \text{ and } 1 \leq j \leq k\}$ and $\Sigma_p = \{;, -, \#, D\} \cup \{V_i \mid 1 \leq i \leq k\}$.

For an edge $e \in E$ between v_l^i and v_s^j where $1 \leq i < j \leq k$ and $1 \leq l, s \leq n$, we write $\mathbf{vt}(e)$ to denote the text $v_l^i - v_s^j$. For $l \in \Sigma_p \cup \Sigma_t$ and $i \in \mathbb{N}$ we write $\mathbf{rp}(l, i)$ to denote the text consisting of repeating the letter l exactly i times. We first define a preliminary text t' as follows.

$$\begin{aligned} & \#l_{1,2}; \mathbf{vt}(e_1^{1,2}); \dots; \mathbf{vt}(e_m^{1,2}); r_{1,2}\# \cdots \#l_{1,k}; \mathbf{vt}(e_1^{1,k}); \dots; \mathbf{vt}(e_m^{1,k}); r_{1,k} \\ & \#l_{2,3}; \mathbf{vt}(e_1^{2,3}); \dots; \mathbf{vt}(e_m^{2,3}); r_{2,3}\# \cdots \#l_{2,k}; \mathbf{vt}(e_1^{2,k}); \dots; \mathbf{vt}(e_m^{2,k}); r_{2,k} \\ & \dots \\ & \#l_{k-1,k}; \mathbf{vt}(e_1^{k-1,k}); \dots; \mathbf{vt}(e_m^{k-1,k}); r_{k-1,k}\# \end{aligned}$$

We also define a preliminary pattern p' as follows.

$$\begin{aligned}
& \#D; V_1 - V_2; D\# \dots \#D; V_1 - V_k; D \\
& \#D; V_2 - V_3; D\# \dots \#D; V_2 - V_k; D \\
& \dots \\
& \#D; V_{k-1} - V_k; D\#
\end{aligned}$$

Let $r = 2(k' + 1)$. Then t is obtained by preceding t' with the text t'' defined as follows.

$$\#; -\mathbf{rp}(+, r)$$

Similarly, p is obtained by preceding p' with the text p'' defined as follows.

$$\#; -\mathbf{rp}(D, r)$$

This completes the construction of t and p . Clearly, t and p can be constructed from G and k in fpt-time (even polynomial time). Furthermore, because $\#\Sigma_p = r + k' = 2(k' + 1) + k' = 3k' + 1$, $|\Sigma_p| = k + 4$, and $\#? = k'$, condition (C1) above is satisfied. To show the remaining condition (C2), we need the following intermediate lemmas.

Lemma 8. *If G has a k -clique, then p k' -GF/GP-matches to t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 1$.*

Proof. Let $\{v_{h_1}^1, \dots, v_{h_k}^k\}$ be the vertices and $\{e_{h_{i,j}}^{i,j} \mid 1 \leq i < j \leq k\}$ be the edges of a k -clique of G with $1 \leq h_j \leq n$ and $1 \leq h_{i,j} \leq m$ for every i and j with $1 \leq i < j \leq k$.

We put k' wildcards on the last k' occurrences of D in p . Informally, these wildcards are mapped in such a way that for every $1 \leq i < j \leq k$ the substring $; V_i - V_j$; of the pattern p is mapped to the substring $; \mathbf{vt}(e_{h_{i,j}}^{i,j});$ of the text t . More formally, for i and j with $1 \leq i < j \leq k$ let $q = (\sum_{o=1}^{o \leq i} (k - o)) + j$. We map the wildcard on the $2(q - 1)$ -th occurrence of the letter D in p' with the text $l_{i,j}; \mathbf{vt}(e_1^{i,j}); \dots; \mathbf{vt}(e_{h_{i,j}-1}^{i,j})$ and similarly we map the wildcard on the $(2(q - 1) + 1)$ -th occurrence of the letter D in p' with the text $\mathbf{vt}(e_{h_{i,j}+1}^{i,j}); \dots; \mathbf{vt}(e_m^{i,j}); r_{i,j}$. Note that in this way every wildcard is mapped to a non-empty substring of t and no two wildcards are mapped to the same substring of t , as required.

We then define the k' -GF/GP-matching function f as follows: $f(;) = ;$, $f(-) = -$, $f(\#) = \#$, $f(V_i) = v_{h_i}^i$, $f(D) = +$, for every i and h_i with $1 \leq i \leq k$ and $1 \leq h_i \leq n$. It is straightforward to check that f together with the mapping for the wildcards maps the pattern p to the text t . \square

Lemma 9. *Let f be a function that k' -GF/GP-matches p to t with $\max_{p \in \Sigma_p} |f(p)| = 1$, then: $f(;) = ;$, $f(-) = -$, $f(\#) = \#$, and $f(D) = +$. Moreover, all wildcards have to be placed on all the k' occurrences of D in p' .*

Proof. We first show that $f(D) = +$. Observe that the string t' is square-free (recall the definition of square-free from Section 2). It follows that every two consecutive occurrences of pattern letters in p'' have to be mapped to a substring of t'' . Because there are $2(k' + 1)$ occurrences of D in p'' it follows that at least two consecutive occurrences of D in p'' are not replaced with wildcards and hence D has to be mapped to a substring of t'' . Furthermore, since all occurrences of D are at the end of p'' , we obtain that D has to be mapped to $+$, as required. Because all occurrences of D in p' have to be mapped to substrings of t' and t' does not contain the letter $+$, it follows that all the k' occurrences of D in p' have to be replaced by wildcards. Since we are only allowed to use at most k' wildcards, this shows the second statement of the lemma. Since no wildcards are used to replace letters in p'' it now easily follows that $f(;) = ;$, $f(-) = -$ and $f(\#) = \#$. \square

Lemma 10. *If p k' -GF/GP-matches to t using a function f with $\max_{p \in \Sigma_p} |f(p)| = 1$, then G has a k -clique.*

Proof. Let f be a function that k' -GF/GP-matches p to t such that $\max_{p \in \Sigma_p} |f(p)| = 1$. We claim that the set $\{f(V_i) \mid 1 \leq i \leq k\}$ is a k -clique of G . Because of Lemma 9, we know that $f(\#) = \#$ and that no occurrence of $\#$ in p is replaced by a wildcard. Since the number of occurrences of $\#$ in t is equal to the number of occurrences of $\#$ in p , we obtain that the i -th occurrence of $\#$ in p is mapped to the i -th

occurrence of $\#$ in t . Consequently, for every i and j with $1 \leq i < j \leq k$, we obtain that the substring $; V_i - V_j$; is mapped to a substring of the string $l_{i,j}; \mathbf{vt}(e_1^{i,j}); \dots; \mathbf{vt}(e_m^{i,j}); r_{i,j}$ in t . Again, using Lemma 9 and the fact that $\max_{p \in \Sigma_p} |f(p)| = 1$, we obtain that both V_i and V_j are mapped to some letter v_l^i and v_s^j for some l and s with $1 \leq l, s \leq n$ such that $\{v_l^i, v_s^j\} \in E$. Hence, $\{f(V_i) \mid 1 \leq i \leq k\}$ is a k -clique of G . \square

Because Condition (C2) is implied by Lemmas 8 and 10, this concludes the proof of Theorem 6.

Theorem 7. *(Max-)GFM and (Max-)GPM are $\mathbf{W}[1]$ -hard parameterized by $\#\Sigma_t$, $\#\Sigma_p$, $|\Sigma_p|$, $\#?$, and $\max |f(?)|$.*

We will show the theorem by a parameterized reduction from MULTICOLORED CLIQUE. Let $G = (V, E)$ be a k -partite graph with partition V_1, \dots, V_k of V . Let $E_{i,j} = \{\{u, v\} \in E \mid u \in V_i \text{ and } v \in V_j\}$ for every i and j with $1 \leq i < j \leq k$. Again, as we stated in the preliminaries we can assume that $|V_i| = n$ and $|E_{i,j}| = m$ for every i and j with $1 \leq i < j \leq k$.

Let $V_i = \{v_1^i, \dots, v_n^i\}$ and $E_{i,j} = \{e_1^{i,j}, \dots, e_m^{i,j}\}$. We construct a text t and a pattern p from G and k such that p GF/GP-matches t if and only if G has a k -clique. The alphabet Σ_t consists of:

- the letter $\#$ (used as a separator);
- one letter a_e for every $e \in E$ (representing the edges of G);
- one letter $\#_i$ for every i with $1 \leq i \leq n$ (used as special separators that group edges from the same vertex);
- the letters $l_{i,j}$, $r_{i,j}$, l_i , r_i for every i and j with $1 \leq i < j \leq k$ (used as dummy letters to ensure injectivity for GPM);
- the letter d_e^v and d^v for every $e \in E$ and $v \in V(G)$ with $v \in e$ (used as dummy letters to ensure injectivity for GPM).

We set $\Sigma_p = \{\#\} \cup \{E_{i,j}, L_{i,j}, R_{i,j}, L_i, R_i, A_i \mid 1 \leq i < j \leq k\} \cup \{D_{i,j} \mid 1 \leq i \leq k \text{ and } 1 \leq j \leq k+1\}$.

For a vertex $v \in V$ and j with $1 \leq j \leq k$ we denote by $E_j(v)$ the set of edges of G that are incident to v and whose other endpoint is in V_j . Furthermore, for a vertex $v \in V(G)$, we write $\mathbf{e}(v)$ to denote the text $\mathbf{el}(v, E_1(v)) \dots \mathbf{el}(v, E_k(v)) d^v$, where $\mathbf{el}(v, E')$, for vertex v and a set E' of edges with $E' = \{e_1, \dots, e_l\}$, is the text $d_{e_1}^v a_{e_1} d_{e_2}^v a_{e_2} \dots d_{e_l}^v a_{e_l}$.

We first define the following preliminary text and pattern strings. Let t_1 be the text:

$$\begin{aligned} & \#l_{1,2}a_{e_1^{1,2}} \dots a_{e_m^{1,2}}r_{1,2}\# \dots \#l_{1,k}a_{e_1^{1,k}} \dots a_{e_m^{1,k}}r_{1,k} \\ & \#l_{2,3}a_{e_1^{2,3}} \dots a_{e_m^{2,3}}r_{2,3}\# \dots \#l_{2,k}a_{e_1^{2,k}} \dots a_{e_m^{2,k}}r_{2,k} \\ & \dots \\ & \#l_{k-1,k}a_{e_1^{k-1,k}} \dots a_{e_m^{k-1,k}}r_{k-1,k} \end{aligned}$$

Let t_2 be the text:

$$\begin{aligned} & \#l_1\#_1\mathbf{e}(v_1^1)\#_1 \dots \#_n\mathbf{e}(v_n^1)\#_nr_1 \\ & \dots \\ & \#l_k\#_1\mathbf{e}(v_1^k)\#_1 \dots \#_n\mathbf{e}(v_n^k)\#_nr_k\# \end{aligned}$$

Let p_1 be the pattern:

$$\begin{aligned} & \#L_{1,2}E_{1,2}R_{1,2}\# \dots \#L_{1,k}E_{1,k}R_{1,k} \\ & \#L_{2,3}E_{2,3}R_{2,3}\# \dots \#L_{2,k}E_{2,k}R_{2,k} \\ & \dots \\ & \#L_{k-1,k}E_{k-1,k}R_{k-1,k} \end{aligned}$$

For i, j with $1 \leq i, j \leq k$, let $I(i, j)$ be the letter $E_{i,j}$ if $i < j$, the letter $E_{j,i}$ if $i > j$ and the empty string if $i = j$. We define $\mathbf{p}(1)$ to be the pattern:

$$A_1 D_{1,2} I(1, 2) D_{1,3} I(1, 3) \cdots D_{1,k} I(1, k) D_{1,k+1} A_1$$

we define $\mathbf{p}(k)$ to be the pattern:

$$A_k D_{k,1} I(k, 1) D_{k,2} I(k, 2) \cdots D_{k,k-1} I(k, k-1) D_{k,k+1} A_k$$

and for every i with $1 < i < k$, we define $\mathbf{p}(i)$ to be the pattern:

$$A_i D_{i,1} I(i, 1) D_{i,2} I(i, 2) \cdots D_{i,i-1} I(i, i-1) \\ D_{i,i+1} I(i, i+1) \cdots D_{i,k} I(i, k) D_{i,k+1} A_i$$

Then p_2 is the pattern:

$$\#L_1 \mathbf{p}(1) R_1 \# \cdots \#L_k \mathbf{p}(k) R_k \#$$

We also define t_0 to be the text $\#\#$ and p_0 to be the pattern $\#\#$. Then, t is the concatenation of t_0 , t_1 and t_2 and p is a concatenation of p_0 , p_1 and p_2 .

This completes the construction of t and p . Clearly, t and p can be constructed from G and k in fpt-time (even polynomial time). Furthermore, $\#\Sigma_t = \binom{k}{2} + k + 3$, $\#\Sigma_p = \binom{k}{2} + k + 3$, $|\Sigma_p| = k(k+1) + 3\binom{k}{2} + 3k + 1$ and hence bounded by k , as required. It remains to show that G has a k -clique if and only if p GF/GP-matches t .

Lemma 11. *If G has a k -clique then p GF/GP-matches t .*

Proof. Let $\{v_{h_1}^1, \dots, v_{h_k}^k\}$ be the vertices and $\{e_{h_i,j}^{i,j} \mid 1 \leq i < j \leq k\}$ be the edges of a k -clique of G with $1 \leq h_i \leq n$ and $1 \leq h_{i,j} \leq m$ for every i and j with $1 \leq i < j \leq k$.

We define the function f that GF/GP-matches p to t as follows: We set $f(\#) = \#$ and $f(;) = ;$. Moreover, for every i and j with $1 \leq i < j \leq k$, we set $f(E_{i,j}) = a_{e_{h_i,j}^{i,j}}$, $f(A_i) = \#_i$, $f(L_{i,j}) = l_{i,j} a_{e_1^{i,j}} \cdots a_{e_{h_{i,j}-1}^{i,j}}$, $f(R_{i,j}) = a_{e_{h_{i,j}+1}^{i,j}} \cdots a_{e_m^{i,j}} r_{i,j}$, $f(L_i) = ; l_i \#_1 \mathbf{e}(v_1^i) \#_1 \cdots \#_{h_i-1} \mathbf{e}(v_{h_i-1}^i) \#_{h_i-1}$, and $f(R_i) = \#_{h_i+1} \mathbf{e}(v_{h_i+1}^i) \#_{h_i+1} \cdots \#_n \mathbf{e}(v_n^i) \#_n r_i$.

For every i and j with $i \neq j$, let $e(i, j)$ be the edge $e_{h_i,j}^{i,j}$ if $i < j$ and the edge $e_{h_j,i}^{j,i}$, otherwise. Then, the letters $D_{i,j}$ are mapped as follows:

- For every i and j with $1 \leq i \leq k$, $2 \leq j \leq k$, $i \neq j$, and $(i, j) \neq (1, 2)$, we map $f(D_{i,j})$ to the substring of $\mathbf{e}(v_{h_i}^i)$ in between the occurrences (and not including these occurrences) of the letters $e(i, j-1)$ and $e(i, j)$.
- We map $f(D_{1,2})$ to be the prefix of $\mathbf{e}(v_{h_1}^1)$ ending before the letter $e(1, 2)$.
- For every i with $2 \leq i \leq k$, we map $f(D_{i,1})$ to the prefix of $\mathbf{e}(v_{h_i}^i)$ ending before the letter $e(i, 1)$.
- For every i with $1 \leq i < k$, we map $f(D_{i,k+1})$ to the suffix of $\mathbf{e}(v_{h_i}^i)$ starting after the letter $e(i, k)$.
- We map $f(D_{k,k+1})$ to be the suffix of $\mathbf{e}(v_{h_1}^1)$ starting after the letter $e(k, k-1)$.

It is straightforward to check that f GF/GP-matches p to t . □

Lemma 12. *If p GF/GP-matches t then G has a k -clique.*

Proof. Let f be a function that GF/GP-matches p to t . We first show that $f(\#) = \#$. Suppose for a contradiction that $f(\#) \neq \#$. Because t and p start with $\#\#$ it follows that $f(\#)$ is a string that starts with $\#\#$. However, t does not contain any other occurrence of the string $\#\#$ and hence the remaining occurrences of $\#$ in p cannot be matched by f .

Because t and p have the same number of occurrences of $\#$, it follows that the i -th occurrences of $\#$ in p has to be mapped to the i -th occurrence of $\#$ in t . We obtain that:

- (1) For every i, j with $1 \leq i < j \leq k$, the substring $L_{i,j}E_{i,j}R_{i,j}$ of p has to be mapped to the substring $l_{i,j}a_{e_1^{i,j}} \cdots a_{e_m^{i,j}}r_{i,j}$ of t .
- (2) For every i with $1 \leq i \leq k$, the substring $L_i \mathbf{p}(i) R_i$ of p has to be mapped to the substring $l_i \#_1 \mathbf{e}(v_1^i) \#_1 \cdots \#_n \mathbf{e}(v_n^i) \#_n r_i$ of t .

Because for every i with $1 \leq i \leq k$ the letters $\#_j$ are the only letters that occur more than once in the substring $l_i \#_1 \mathbf{e}(v_1^i) \#_1 \cdots \#_n \mathbf{e}(v_n^i) \#_n r_i$ of t , we obtain from (2) that A_i has to be mapped to $\#_j$ for some j with $1 \leq j \leq n$. Consequently:

- (3) for every i with $1 \leq i \leq k$, the substring $\mathbf{p}(i)$ of p has to be mapped to a substring $\#_j \mathbf{e}(v_j^i) \#_j$ of t for some j with $1 \leq j \leq n$.

It follows from (1) that for every i, j with $1 \leq i < j \leq k$, $f(E_{i,j})$ is mapped to some edges between V_i and V_j . W.l.o.g. we can assume that $f(E_{i,j})$ is mapped to exactly one edge between V_i and V_j , because mapping it to more than one edge would make the matching of the latter occurrences of $E_{i,j}$ in p even harder, i.e., whenever a latter occurrence of $E_{i,j}$ in p can be mapped to more than one edge it also can be mapped to any of these edges. Furthermore, because of (3) it follows that for every i with $1 \leq i \leq k$, it holds that the edges mapped to any $E_{l,r}$ with $1 \leq l < r \leq k$ such that $l = i$ or $r = i$ have the same endpoint in V_i . Hence, the set of edges mapped to the letters $E_{i,j}$ for $1 \leq i < j \leq k$ form a k -clique of G . \square

This concludes the proof of Theorem 7.

References

- [1] Amihood Amir and Igor Nor. Generalized function matching. *Journal of Discrete Algorithms*, 5(3):514–523, 2007.
- [2] Dana Angluin. Finding patterns common to a set of strings (extended abstract). In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 130–141, 1979.
- [3] Dana Angluin. Finding patterns common to a set of strings. *Journal of Computer and System Sciences*, 21(1):46 – 62, 1980.
- [4] Raphaël Clifford, Aram W. Harrow, Alexandru Popa, and Benjamin Sach. Generalised matching. In Jussi Karlgren, Jorma Tarhio, and Heikki Hyvärinen, editors, *String Processing and Information Retrieval, 16th International Symposium, SPIRE 2009, Saariselkä, Finland, August 25-27, 2009, Proceedings*, volume 5721 of *Lecture Notes in Computer Science*, pages 295–301. Springer, 2009.
- [5] Raphaël Clifford and Alexandru Popa. (In)approximability results for pattern matching problems. In Jan Holub and Jan Zdárek, editors, *Proceedings of the Prague Stringology Conference 2010, Prague, Czech Republic, August 30 - September 1, 2010*, pages 52–62. Prague Stringology Club, Department of Theoretical Computer Science, Faculty of Information Technology, Czech Technical University in Prague, 2010.
- [6] Rodney G. Downey and Michael R. Fellows. *Parameterized Complexity*. Monographs in Computer Science. Springer Verlag, New York, 1999.
- [7] Andrzej Ehrenfeucht and Grzegorz Rozenberg. Finding a homomorphism between two words in np-complete. *Information Processing Letters*, 9(2):86 – 88, 1979.
- [8] Henning Fernau and Markus L. Schmid. Pattern matching with variables: A multivariate complexity analysis. In Johannes Fischer and Peter Sanders, editors, *Combinatorial Pattern Matching, 24th Annual Symposium, CPM 2013, Bad Herrenalb, Germany, June 17-19, 2013. Proceedings*, volume 7922 of *Lecture Notes in Computer Science*, pages 83–94. Springer, 2013.

- [9] Henning Fernau, Markus L. Schmid, and Yngve Villanger. On the parameterised complexity of string morphism problems. In Anil Seth and Nisheeth K. Vishnoi, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2013, December 12-14, 2013, Guwahati, India*, volume 24 of *LIPICs*, pages 55–66. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2013.
- [10] Jörg Flum and Martin Grohe. *Parameterized Complexity Theory*, volume XIV of *Texts in Theoretical Computer Science. An EATCS Series*. Springer Verlag, Berlin, 2006.
- [11] Dominik D. Freydenberger, Daniel Reidenbach, and Johannes C. Schneider. Unambiguous morphic images of strings. In Clelia Felice and Antonio Restivo, editors, *Developments in Language Theory*, volume 3572 of *Lecture Notes in Computer Science*, pages 248–259. Springer Berlin Heidelberg, 2005.
- [12] Tao Jiang, Efim Kinber, Arto Salomaa, Kai Salomaa, and Sheng Yu. Pattern languages with and without erasing. *International Journal of Computer Mathematics*, 50(3-4):147–163, 1994.
- [13] Alexandru Mateescu and Arto Salomaa. Finite degrees of ambiguity in pattern languages. *Informatique Théorique et Applications*, 28(3-4):233–253, 1994.
- [14] Yen Kaow Ng and Takeshi Shinohara. Developments from enquiries into the learnability of the pattern languages from positive data. *Theoretical Computer Science*, 397(13):150 – 165, 2008. Forty Years of Inductive Inference: Dedicated to the 60th Birthday of Rolf Wiehagen.
- [15] Krzysztof Pietrzak. On the parameterized complexity of the fixed alphabet shortest common supersequence and longest common subsequence problems. *Journal of Computer and System Sciences*, 67(4):757–771, 2003.
- [16] Daniel Reidenbach. A non-learnable class of e-pattern languages. *Theoretical Computer Science*, 350(1):91 – 102, 2006.
- [17] Daniel Reidenbach. Discontinuities in pattern inference. *Theoretical Computer Science*, 397(13):166 – 193, 2008. Forty Years of Inductive Inference: Dedicated to the 60th Birthday of Rolf Wiehagen.
- [18] Markus L. Schmid. A note on the complexity of matching patterns with variables. *Information Processing Letters*, 113(19–21):729–733, 2013.
- [19] Takeshi Shinohara. Polynomial time inference of extended regular pattern languages. In Eiichi Goto, Koichi Furukawa, Reiji Nakajima, Ikuo Nakata, and Akinori Yonezawa, editors, *RIMS Symposia on Software Science and Engineering*, volume 147 of *Lecture Notes in Computer Science*, pages 115–127. Springer Berlin Heidelberg, 1983.